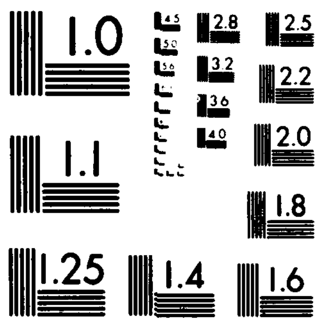
AD-A114 579 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
BAYES ESTIMATION OF A MULTIVARIATE DENSITY.(U)
FEB 82 T LEONARD DAAG29-80-C-0041
UNCLASSIFIED MRC-TSR-2342 NL

END
DATE
FILMED
6 82
DTIC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS 1963-A

MRC Technical Summary Report #2342

BAYES ESTIMATION OF A MULTIVARIATE DENSITY

Tom Leonard

Mathematics Research Center
University of Wisconsin–Madison
610 Walnut Street
Madison, Wisconsin 53706

February 1982

Received January 20, 1982

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

BAYES ESTIMATION OF A MULTIVARIATE DENSITY

Tom Leonard

Technical Summary Report #2342
February 1982

ABSTRACT

The problem addressed concerns the estimation of a p-dimensional multivariate density, given only a set of n observation vectors, together with information that the density function is likely to be reasonably smooth. A solution is proposed which employs up to $n + \frac{1}{2} p(p+1)$ smoothing parameters, all of which may be estimated by their posterior means. This avoids the well-known difficulties, associated with even one-dimensional kernel estimators, of estimating the bandwidth or smoothing parameter by a mathematical procedure. The posterior mean value function, unconditional upon the smoothing parameters, turns out to be a data-based mixture of multivariate t-distributions. The corresponding estimate of the sampling covariance matrix may be viewed as a shrinkage estimator of the Bayes-Stein type. The results involve some finite series which may be evaluated by a straightforward simulation procedure.

AMS(MOS) Subject Classifications: 62G05, 62H12

Key Words: multivariate density estimation, Bayes, Dirichlet, Wishart, multivariate t-distribution, smoothing, bandwidth, covariance matrix, Bayes-Stein, kernel estimator, computer simulation.

Work Unit No. 4 - Statistics and Probability

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

SIGNIFICANCE AND EXPLANATION

A method is proposed for the smooth estimation of a multivariate density, given a finite number of observation vectors. The main procedure suggested employs a generalized kernel estimator, where the nodes are selected via a preliminary cluster analysis. An exchangeable Dirichlet prior for the coefficients is used together with an inverted Wishart prior for the smoothing parameters. The posterior means of these quantities are obtained, and the unconditional posterior mean value function of the multivariate density is shown to be a finite mixture of multivariate t-densities. All results involve the summation of a finite series; this is best evaluated by a simulation technique.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

BAYES ESTIMATION OF A MULTIVARIATE DENSITY

Tom Leonard

1. PRIOR ASSUMPTIONS

Suppose that the observation vectors $\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_n$ are independent, given their common p-dimensional multivariate density $g(\underset{\sim}{x})$; $\underset{\sim}{x} \in R^p$. Generalized kernel estimators of the form

$$g^*(\underset{\sim}{x}) = \sum_{j=1}^{r} \theta_j \phi_j(\underset{\sim}{x}) \qquad \text{for } \underset{\sim}{x} \in R^p; \quad \sum_{j=1}^{r} \theta_j = 1 \tag{1.1}$$

are considered, where

$$\phi_j(\underset{\sim}{x}) = (2\pi)^{-\frac{1}{2}p} |\underset{\sim}{C}|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\xi}_j)^T \underset{\sim}{C}^{-1} (\underset{\sim}{x}-\underset{\sim}{\xi}_j) \quad (j=1,\ldots,r) \tag{1.2}$$

The assumptions in (1.1) and (1.2) place multivariate normal kernels over the r points $\underset{\sim}{\xi}_1,\ldots,\underset{\sim}{\xi}_r$, and then estimate g by a weighted average of these kernels, with weights $\theta_1,\ldots,\theta_r$. The matrix $\underset{\sim}{C}^{-1}$ is related to the idea of bandwidth for one-dimensional kernel estimators; its specification has a large effect upon the estimation of g. (e.g. Silverman, 1978). The unequal weights $\theta_1,\ldots,\theta_r$, for the points $\underset{\sim}{\xi}_1,\ldots,\underset{\sim}{\xi}_r$, contrast the equal weights stipulated by kernel estimators, (for the n kernels centered on the data points $\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_n$). They are designed to avoid the typical practical difficulties associated with equal weights e.g. either oversmoothing or overbumpy tails depending on the choice of bandwidth.

A number of further assumptions are made about $\theta_1,\ldots,\theta_r$; $\underset{\sim}{C}$, and $\underset{\sim}{\xi}_1,\ldots,\underset{\sim}{\xi}_r$. They are

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

(a) The proportions $\theta_1,\ldots,\theta_r$ are taken to possess a Dirichlet prior distribution

$$\pi(\underset{\sim}{\theta}|\alpha,\underset{\sim}{\eta}) \propto \theta_1^{\alpha\eta_1-1}\theta_2^{\alpha\eta_2-1}\ \ldots\ \theta_r^{\alpha\eta_r-1} \tag{1.3}$$

$$(0 < \alpha < \infty;\ \Sigma\eta_j=1)$$

over their simplex of permissible values; here $\eta_1,\ldots,\eta_r$ denote the respective prior means of $\theta_1,\ldots,\theta_r$, and α controls the degree of shrinkage of our posterior estimates towards $\eta_1,\ldots,\eta_r$.

The following procedures are available for choosing $\eta_1,\ldots,\eta_r$ and either choosing or empirically estimating α.

(i) If we possess prior ignorance about $\theta_1,\ldots,\theta_r$ we could set $\alpha = r;\ \eta_1,\ldots,\eta_r = r^{-1}$, yielding a uniform, but proper, prior distribution over the simplex.

(ii) If we regard $\theta_1,\ldots,\theta_r$ as exchangeable then we could set $\eta_1 = \ldots = \eta_r = r^{-1}$ and then estimate α from the data by either a hierarchical or empirical Bayes method (see section 3). This yields data based Bayes - Stein shrinkage estimators for $\theta_1,\ldots,\theta_r$; it is our main suggestion. The shrinkage estimators avoid over concentration of the density estimates at particular nodes ξ_j.

(iii) If we possess prior information to suggest suitable weights for the r kernels, then we could choose $\eta_1,\ldots,\eta_r$ accordingly, and then estimate α from the data, as in (ii).

(b) The matrix $\underset{\sim}{C}$ is taken to possess an inverted Wishart prior distribution

$$\pi(\underset{\sim}{C}|\nu,\underset{\sim}{T}) = c|T|^{\frac{1}{2}\nu}|\underset{\sim}{C}|^{-\frac{1}{2}(\nu+p+1)}\exp\{-\tfrac{1}{2}\text{trace } \underset{\sim}{C}\underset{\sim}{T}\} \qquad (1.4)$$

where

$$c^{-1} = 2^{\nu p/2}\, \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma\left(\frac{\nu-j+1}{2}\right) \qquad (1.5)$$

given its parametric matrix $\underset{\sim}{T}$ and degrees of freedom ν. Here $(\nu-2)^{-1}\underset{\sim}{T}$ is a prior estimate of $\underset{\sim}{C}$ and ν is a prior 'sample size' measuring the strength of this information.

The following choices are available for ν and $\underset{\sim}{T}$.

(i) In situations where there is no prior information about $\underset{\sim}{C}$; $\nu = 1$ and $\underset{\sim}{T}$ equailing the $n \times n$ identity matrix are reasonable choices.

(ii) If there is related information from other samples then this might also be represented by the specification of values for ν and $\underset{\sim}{T}$.

(iii) Possible extensions to the present analysis include empirical estimation of ν when T is specified. Also $\underset{\sim}{T}$ could be assumed to take the intraclass form; it is then possible to empirically estimate ν together withe the variance and correlation appearing in the special covariance structure.

(c) Possible empirical choices of $\underset{\sim}{\xi}_1,\ldots,\underset{\sim}{\xi}_r$ involve

(i) Setting $r = n$ and $\underset{\sim}{\xi}_j = \underset{\sim}{x}_j$ for $j = 1,\ldots,n$; a frequent choice in the literature of kernel estimation. We, of course, make this substitution after the prior to posterior analysis has been performed for fixed $\xi_1,\ldots,\xi_r$. However, when $r = n$, our procedure will often undersmooth; therefore choices (ii) and (iii) will usually be preferable.

(ii) Arranging $\underset{\sim}{\xi}_1,\ldots,\underset{\sim}{\xi}_r$ to lie on an equally spaced p-dimensional lattice; the choice of r depending upon the fineness of the grid.

(iii) Performing a preliminary cluster analysis, and then putting at least one $\underset{\sim}{\xi}$ in the center of each cluster.

(iv) In both (ii) and (iii) suitable values of r may be obtained by comparing realizations of the prior predictive density, and maximizing with respect to r (see section 6).

Of the above possibilities the choice a(ii), b(i) and c(iii) seems to be reasonable if the objective is to carry out a data-based analysis, which is virtually free from choice of prior parameter values.

Note that our prior assumptions are very different from those proposed by Leonard (1973, 1978, 1982) who uses a Gaussian prior/logistic transform approach on function space. Whilst the covariances of Gaussian processes are useful in one-dimension, they lead to overwhelmingly complicated computations in several dimensions so that it is necessary to decrease the complexity of the prior structure in order to obtain reasonable results.

Note that in case c(iii) the procedure we describe is equivalent in spirit to the data analytic procedure of assigning proportions to each cluster in accordance with the number of observations in the cluster, and then estimating $\underset{\sim}{C}$ as the common parameter matrix in the corresponding discrete mixture of multivariate normal distributions.

2. ESTIMATION OF $\underset{\sim}{\theta}$

The likelihood of $\underset{\sim}{\theta}$ and $\underset{\sim}{C}$, given the assumption in (1.2), and the n observation vectors, is denoted by

$$\ell(\underset{\sim}{\theta},\underset{\sim}{C}|\underset{\sim}{x}^{(n)}) \propto \prod_{i=1}^{n} \sum_{j=1}^{r} \theta_j \phi_j(\underset{\sim}{x}_i)$$
$$\propto \sum_{\Omega} \prod_{i=1}^{n} \theta_{\omega(i)} \phi_{\omega(i)}(\underset{\sim}{x}_i) \tag{2.1}$$

where Ω is the space of all possible $n \times 1$ vectors $\omega = (\omega(1),\ldots,\omega(n))$, where, for $i = 1,\ldots,n$, $\omega(i)$ is a mapping $\omega(i)$: $i \to (1,\ldots,r)$ from i to the set of the first r integers. Under the choices in (1.2) for the $\phi_j(\underset{\sim}{x})$, (2.1) implies that

$$\ell(\underset{\sim}{\theta},\underset{\sim}{C}|\underset{\sim}{x}^{(n)}) \propto |\underset{\sim}{C}|^{-\frac{1}{2}n} \sum_{\Omega} \prod_{j=1}^{r} \theta_j^{n_j} \exp\{-\tfrac{1}{2}\text{trace } \underset{\sim}{C}^{-1} \underset{\sim}{D}_{\omega}\} \tag{2.2}$$

where

$$n_j = {}^{\#}[i\varepsilon(1,\ldots,n):\omega(i)=j] \tag{2.3}$$

with $\Sigma n_j = n$, and

$$\underset{\sim}{D}_{\omega} = \sum_{i=1}^{m} (\underset{\sim}{x}_i-\underset{\sim}{\xi}_{\omega(i)})(\underset{\sim}{x}_i-\underset{\sim}{\xi}_{\omega(i)})^T \tag{2.4}$$

Note that Ω contains r^n elements; however summations with respect to $\omega\epsilon\Omega$ may be evaluated by computer simulations generating random elements from Ω. This procedure is straightforward, the same set of vectors simulated for $\underset{\sim}{\omega}$ may be used for each of the summations involved in this analysis.

The joint posterior distribution of $\underset{\sim}{\theta}$ and $\underset{\sim}{C}$ is now

$$\pi(\underset{\sim}{\theta},\underset{\sim}{C}|\underset{\sim}{x}^{(n)},\alpha,n,\nu,T)$$

$$\propto \pi(\underset{\sim}{\theta}|\alpha,n)\ \pi|\underset{\sim}{C}|\nu,\underset{\sim}{T})\ \ell(\underset{\sim}{\theta},\underset{\sim}{C}|\underset{\sim}{x}^{(n)})$$

$$\propto |\underset{\sim}{C}|^{-\frac{1}{2}(\nu+p+n-1)} \sum_{\Omega} \prod_{j=1}^{r} \theta_j^{\alpha_j+n_j-1} \exp\{-\tfrac{1}{2}\mathrm{trace}[\underset{\sim}{C}^{-1}(\underset{\sim}{T}+\underset{\sim}{D}_{\omega})]\} \qquad (2.5)$$

Integrating with respect to $\underset{\sim}{C}$, we find that the posterior density of $\underset{\sim}{\theta}$ is a mixture of Dirichlet densities,

$$\pi(\underset{\sim}{\theta}|\underset{\sim}{x}^{(n)},\alpha,n) = \sum_{\Omega} A_{\omega}\delta(\underset{\sim}{\theta}|\underset{\sim}{n})/\sum_{\Omega} A_{\omega} \qquad (2.6)$$

where

$$\delta(\underset{\sim}{\theta}) = k_{\omega}(\underset{\sim}{n}) \prod_{j=1}^{r} \theta_j^{\alpha_j+n_j-1} \qquad (2.7)$$

with

$$k_{\omega}(\underset{\sim}{n}) = \Gamma(\alpha+n)/\prod_{j} \Gamma(\alpha_j+n_j) \qquad (2.8)$$

and

$$A_\omega = |\underset{\sim}{T}+\underset{\sim}{D}_\omega|^{-\frac{1}{2}(\nu+n)}/k_\omega(n) \tag{2.9}$$

Bayes estimates may be readily elicited from (2.6), for example, the posterior mean vector of $\underset{\sim}{\theta}$, conditional upon α and $\underset{\sim}{n}$ is given by

$$E(\underset{\sim}{\theta}|\underset{\sim}{x}^{(n)},\alpha,\underset{\sim}{n}) = \sum_{\Omega} A_\omega \underset{\sim}{a}_\omega / \sum_{\Omega} A_\omega \tag{2.10}$$

where

$$\underset{\sim}{a}_\omega = (\alpha_1+n_1,\ldots,\alpha_r+n_r)^T/(\alpha+n) \tag{2.11}$$

Higher moments may be obtained similarly by averaging the higher moments of the Dirichlet distribution in (2.7).

3. EMPIRICAL ESTIMATION OF THE SHRINKAGE PARAMETER

Integrating the joint density in (2.5) with respect to both $\underset{\sim}{\theta}$ and $\underset{\sim}{C}$, gives, with the judicious inclusion of an extra factor in the proportionality constant of (2.5),

$$\iint \pi(\underset{\sim}{\theta},\underset{\sim}{C}|\underset{\sim}{x}^{(n)},\alpha,n,\nu,\underset{\sim}{T})\,d\underset{\sim}{\theta}\,d\underset{\sim}{C}$$

$$\propto \sum_{\Omega} A_\omega B_\omega(\alpha) \tag{2.12}$$

where

$$B_\omega(\alpha) = \frac{\Gamma(\alpha)\prod_j \Gamma(\alpha\xi_j+n_j)}{\Gamma(\alpha+n)\prod_j \Gamma(\alpha\xi_j)} \tag{2.13}$$

and A_ω is defined in (2.9).

The expression in (2.12) is, as a function of α, proportional to the marginal density of $\underset{\sim}{x}^{(n)}$, given α. It may therefore be regarded as proportional to the "marginal likelihood" of α.

The posterior density of α is therefore given by,

$$\pi(\alpha|\underset{\sim}{x}^{(n)},\eta) \propto \pi(\alpha) \sum_\Omega A_\omega B_\omega(\alpha) \qquad (0 < \alpha < \infty) \tag{2.14}$$

where $\pi(\alpha)$ represents the prior density of α. It is then possible, in principle, to calculate the posterior mean of $\underset{\sim}{\theta}$, unconditional upon α, from

$$E(\underset{\sim}{\theta}|\underset{\sim}{x}^{(n)},\underset{\sim}{\eta}) = \int E(\underset{\sim}{\theta}|\underset{\sim}{x}^{(n)},\alpha,\underset{\sim}{\eta})\pi(\alpha|\underset{\sim}{x}^{(n)},\underset{\sim}{\eta})d\alpha \tag{2.15}$$

where the contributions to the integrand of (2.15) are given in (2.10) and (2.4) respectively.

Some approximations lead to simpler computations. Leonard (1977) considers posterior densities of the form

$$\pi^*(\alpha|x^{(n)}) \propto \pi(\alpha)\sum_\Omega A_\omega B_\omega(\alpha) \tag{2.16}$$

and shows that, with $\tau = (1+\alpha)/(n+\alpha)$ and

$$X^2 = \sum_i (n_i - n\eta_i)^2 / n\eta_i \tag{2.17}$$

the posterior mean of τ is, for large r, approximated by

$$\tau^* = \max\left(\frac{r-1}{X^2}, 1\right). \tag{2.18}$$

Therefore, as an approximation, the corresponding values for $\alpha = (n\tau-1)/(1-\tau)$ may be substituted for α on the right hand side of (2.10). This works even when $\alpha = \infty$ since then $\alpha_j/\alpha = \eta_j$. More precise versions of τ^*, in terms of incomplete Gamma functions, are available.

4. ESTIMATION OF $\underset{\sim}{C}$.

Integrating the joint density in (2.5) with respect to $\underset{\sim}{\theta}$ we find that the posterior density of $\underset{\sim}{C}$ is

$$\pi(\underset{\sim}{C}|\underset{\sim}{x}) \propto |\underset{\sim}{C}|^{-\frac{1}{2}(\nu+p+n-1)} \sum_{\Omega} u(\omega)\exp\{-\tfrac{1}{2}\underset{\sim}{C}^{-1}(T+D_\omega)\} \tag{4.1}$$

where

$$u(\omega) = \prod_j \Gamma(\alpha_j+n_j)/\Gamma(\alpha+n). \tag{4.2}$$

The density in (4.1) is a discrete mixture of Wishart densities. The posterior mean matrix of $\underset{\sim}{C}$ is

$$E(\underset{\sim}{C}|\underset{\sim}{x}) = (\nu+n-2)^{-1} \sum_{\Omega} u(\omega)(\underset{\sim}{T}+\underset{\sim}{D}_{\omega})/\sum_{\Omega} u(\omega)$$

$$= (\nu+n-2)^{-1}\underset{\sim}{T} + (\nu+n-2)^{-1} \sum_{\Omega} u(\omega)\underset{\sim}{D}_{\omega}/\sum_{\Omega} u(\omega) \qquad (4.3)$$

Here $(\nu-2)^{-1}\underset{\sim}{T}$ is the prior estimate and the second term on the right handside of (4.3) averages the matrix in (2.4).

5. POSTERIOR MEAN VALUE FUNCTION OF THE MULTIVARIATE DENSITY

The posterior mean of $g^*(\underset{\sim}{x})$ in (1.1), conditional upon the hyperparameters α, η, ν, and $\underset{\sim}{T}$ is

$$E(g^*(\underset{\sim}{x})|\underset{\sim}{x}^{(n)}) = \sum_{j=1}^{r} E(\theta_j\phi_j(\underset{\sim}{x})|\underset{\sim}{x}^{(n)})$$

where

$$E(\theta_\ell\phi_\ell(\underset{\sim}{x})|x^{(n)}) = E(\theta_\ell\frac{1}{(2\pi)^{\frac{1}{2}r}|\underset{\sim}{C}|^{\frac{1}{2}}} \exp\{-\tfrac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\xi}_\ell)^T\underset{\sim}{C}^{-1}(\underset{\sim}{x}-\underset{\sim}{\xi}_\ell)|x^{(n)})$$

where the expectation is with respect to the joint distribution in (2.5). After some manipulation we find that

$$E(g^*(\underset{\sim}{x})|\underset{\sim}{x}^{(n)})$$

$$\propto \sum_{\Omega} \sum_{\ell=1}^{r} \frac{\alpha_\ell+n_\ell}{\alpha+n} u(\omega)|\underset{\sim}{T} + \underset{\sim}{D}_\omega + (\underset{\sim}{x}-\underset{\sim}{\xi}_\ell)(\underset{\sim}{x}-\underset{\sim}{\xi}_\ell)^T|^{-\frac{1}{2}(\nu+n-1)} \quad (\underset{\sim}{x}\epsilon R^n) \qquad (5.1)$$

with $u(\omega)$ defined in (4.2).

The weighted average of multivariate t-densities in (5.1) is a Bayes estimate of $g^*(\underset{\sim}{x})$ under both component wise and integrated squared error loss; it is also the predictive density for a further observation vector $\underset{\sim}{x}$.

6. MISCELLANEOUS TOPICS

The posterior means of the central moments correspond to the central moments of the distribution in (5.1). They may however be obtained more directly from (1.1) and the results in sections 2-4. For example, the mean vector and covariance matrix are estimated by

$$\underset{\sim}{\mu} = \sum_{j=1}^{r} \xi_j E(\theta_j | \underset{\sim}{x}^{(n)})$$

and

$$\underset{\sim}{\Sigma} = \sum_{j=1}^{r} [E(\underset{\sim}{\theta}_j \underset{\sim}{C} | \underset{\sim}{x}^{(n)}) + \underset{\sim}{\xi}_j \underset{\sim}{\xi}_j^T E(\underset{\sim}{\theta}_j | \underset{\sim}{x}^{(n)})] - \underset{\sim}{\mu}\underset{\sim}{\mu}^T .$$

The covariance matrix $\underset{\sim}{\Sigma}$ may be compared with the sample covariance matrix calculated from $\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n$. It will smooth the sample covariance matrix by shrinking its elements to allow for our various sampling and prior assumptions. When both α and $\underset{\sim}{C}$ are estimated empirically $\underset{\sim}{\Sigma}$ will take the form of a Bayes-Stein estimate because the amounts of shrinkage depend primarily upon the data.

Whilst the choice $r = n$ is sometimes possible, the statistician might prefer to work with a fewer number of ξ's e.g. with $p = 2$ he might, for n around 200, spread the ξ's over a 10×10 grid. In such circumstances, we may compare different values of r by the criterion

$$\log p(\underset{\sim}{x}^{(n)}|\alpha,r) = \log \sum_{\Omega} A_{\omega}B_{\omega}(\alpha)-\tfrac{1}{2}rn \log \pi \qquad (6.1)$$

where A_{ω} and B_{ω} are defined in (2.9) and (2.13).

The expression in (6.1) is the log of the (prior predictive) marginal density of $x^{(n)}$, given r and the hyperparameters. If several different values of r possess equal prior probabilities, then maximizing (6.1) yields the value of r with maximum posterior probability.

Most of the results in this paper have been stated explicitly in terms of one-dimensional finite series. This might be viewed as surprising in light of the apparent complexity of the problem of multivariate density estimation. The evaluation of the finite series via computer simulations should not cause any untoward difficulites. Whilst elements $\underset{\sim}{\omega}$ of $\underset{\sim}{\Omega}$ should, formally speaking, be chosen at random without replacement, the chance of choosing the same $\underset{\sim}{\omega}$ twice is infinitesmally small, therefore the simulations may quite reasonable proceed with replacement.

REFERENCES

LEONARD, T. (1973) A Bayesian method for histograms. Biometrika 60, 297-308.

LEONARD, T. (1977) A Bayesian approach to some multinomial and pre-testing problems. J.A.S.A. 72, 869-874.

LEONARD, T. (1978) Density Estimation, Stochastic Processes, and Prior Information (with Discussion). J. Roy. Statist. Soc., B. 40, 113-146.

LEONARD, T. (1982) The empirical Bayesian estimation of a one-dimensional function. University of Wisconsin-Madison, Mathematics Research Center. Technical Report.

SILVERMAN, B. W. (1978) Choosing the window width when estimating a density. Biometrika, 65, 1-13.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 2342	2. GOVT ACCESSION NO. AD-A114 579	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) BAYES ESTIMATION OF A MULTIVARIATE DENSITY		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Tom Leonard		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 4 - Statistics & Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE February 1982 13. NUMBER OF PAGES 13
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Multivariate density estimation, Bayes, Dirichlet, Wishart, multivariate t-distribution, smoothing, bandwidth, covariance matrix, Bayes-Stein, kernel estimator, computer simulation		

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The problem addressed concerns the estimation of a p-dimensional multivariate density, given only a set of n observation vectors, together with information that the density function is likely to be reasonably smooth. A solution is proposed which employs up to $n + \frac{1}{2}p(p+1)$ smoothing parameters, all of which may be estimated by their posterior means. This avoids the well-known difficulties, associated with even one-dimensional kernel estimators, of estimating the bandwidth or smoothing parameter by a mathematical procedure. The posterior mean value function, unconditional upon the smoothing parameters, next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract (continued)

turns out to be a data-based mixture of multivariate t-distributions. The corresponding estimate of the sampling covariance matrix may be viewed as a shrinkage estimator of the Bayes-Stein type. The results involve some finite series which may be evaluated by a straightforward simulation procedure.

DATE
ILME
-8